

15 STATISTICS and PROBABILITY

D'HONT RULE

D'Hondt method is a rule for allocating seats in the Parliament. A Belgian mathematician, Victor D'Hondt, was its creator.

D'Hondt slightly favours large parties and coalitions over dispersed small parties.

Legislatures using this system include many countries, Spain is one of them.

You can find the Spanish law for the elections in BOE: 20 jun 1985. Art 1463. Organic law 19 jun 1985. Elections.

Art 163:

1. The allocation of seats will have the following rules:

- Not taken into account those candidates who have not obtained at least 3% of the valid votes cast.
- We order from highest to lowest in a column the numbers of votes obtained by the other candidates.
- Divide the number of votes obtained by each candidate for 1, 2, 3 etc., up to equal the number of seats.

The seats are allocated to candidates who obtain higher ratios in the table, attending a decreasing order.

Example: 480,000 valid votes. Eight deputies. 6 nominations

Nominations have obtained the

following votes:

- A: 168,000 votes.
- B: 104,000 votes.
- C: 72,000 votes.
- D: 64,000 votes.
- E: 40,000 votes.
- F: 32,000 votes.



Division by	1	2	3	4	5	6	7	8
A	168000							
B	104000							
C	72000							
D	64000							
E	40000							
F	32000							

You have to complete the chart and according to the rule decide the 6 elected candidates.

1. FREQUENCY TABLE

Statistics

Statistics is the science that studies the methods to collect meaningful data, to organize them and to analyze them.

The name comes from the word State because the origin of this science was the study of populations.

The picture shows the population distribution in a country according to the age: the three groups are: from 0-16 years; 17-65 years and 66 and up.

Could you write next to the sector the approximate percentage corresponding to each group?



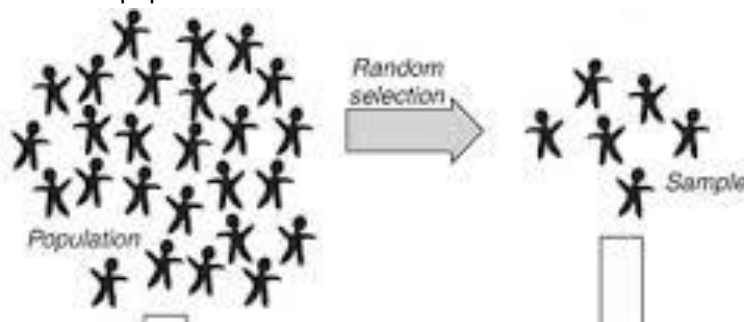
Definitions

Population

A population in Statistics is the set of all elements that are involved in a statistical study.

Sample

A sample is a representative set of the population. The number of individuals in a sample is less than the population.



Write an example of population and sample. For example, a population could be the school and a sample the representative students.

Statistical character

It is a feature or property we study about a population.

There are qualitative characters or quantitative characters.

Qualitative characters can be categorical - Ordination is not possible- like the colour, the brand of a car,... or ordinal -it is possible- like bad, normal, good.

Numerical characters have a division into: discrete -the possible values are isolated- like the number of siblings or continuous -the possible value are classified by intervals- like the height, the weight,...

We represent characters using variables; this is the statistical variable. The variable could be the brand, the opinion, the height,...

Frequency table



The frequency of a value in a statistical variable is a measurement of the number of times that this value appears.

Absolute frequency is the number of times that a certain value appears: n_i

Relative frequency is the ratio between the absolute frequency and the total number of data: f_i

Percentage frequency is the frequency over 100: %

Complete the chart according to the above population

Variable	n_i	f_i	%
			
			
Totales			

2. CHARTS

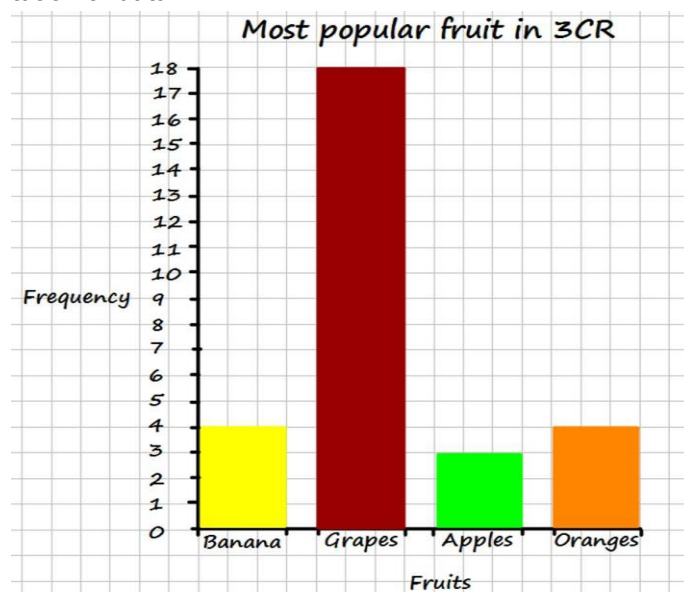
A chart is a graphical representation of data.

BAR CHART

We use a bar for each data with a proportional length to its frequency.

The chart shows the chosen fruit in a classroom.

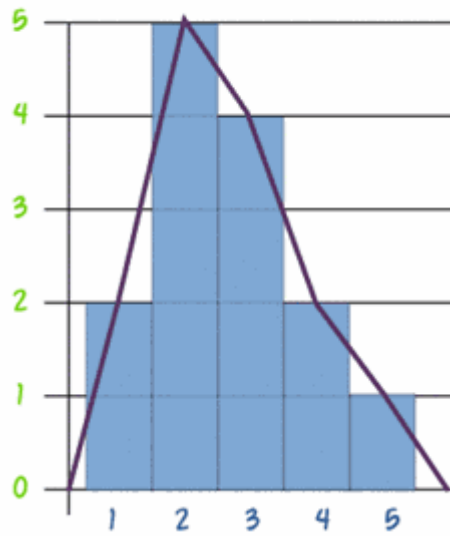
How many pupils are there in the classroom?



FREQUENCY POLYGONS

It is the resulting polygon we get by joining the tips of the bars orderly. This chart only has sense for numerical or ordinal variables.

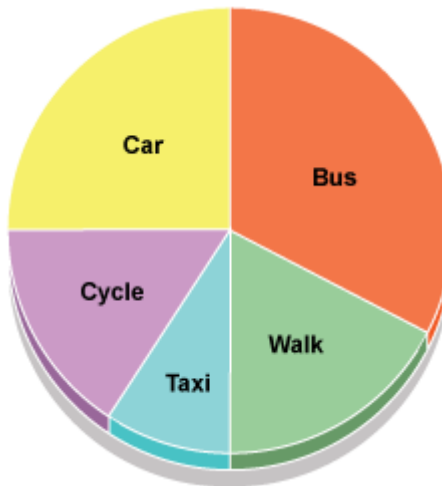
Scores: 1,1,2,2,2,2,2,3,3,3,3,4,4,5



PIE CHART

It is the result of assigning a proportional circular sector to each data till completing the circumference.

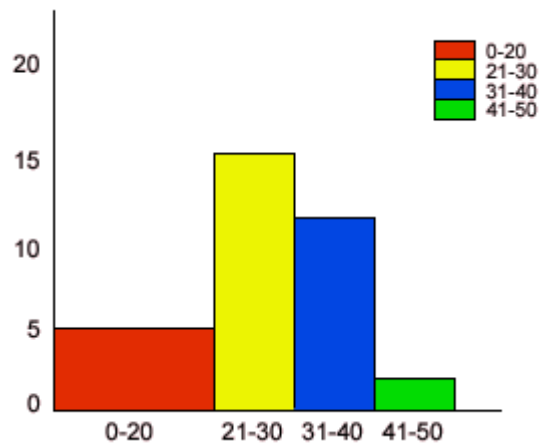
What is the percentage for each kind of transport?



HISTOGRAM

It is useful to represent continuous variables.

First we divide the range of the variable into intervals. Then, we assign a proportional rectangle to the frequency of each interval.



3. MEASURES OF CENTRAL TENDENCY

Statistics use different values obtained from the data in order to get conclusions from the data.

Some of them are the central tendency. They inform us about the central values of the variable in study. They are useful as representative values of the population.

AVERAGE

It is the most popular. It is the result to divide the sum of the data into the number of data:

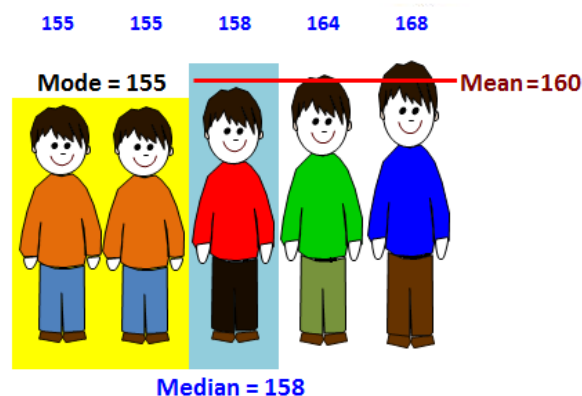
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MODE

It is the most repeated value in the population.

MEDIAN

It is the value located in the center of all the ordering values. That is, median would be greater than the half and lower than the other half.



Check the mean is 160.

PROBLEMS AND EXERCISES

INSTITUTO NACIONAL DE ESTADÍSTICA

Si España fuera un pueblo de 100 habitantes

Subtitulado en inglés

Prueba tus conocimientos

Estadísticas territoriales

1. **FREQUENCY TABLE**

1. The weights of eight children are the following: 14, 32, 18, 25, 40, 42, 35 and 30 Kg. Calculate the average weight.
2. From the study of a group of forty married couples by the number of siblings has the next result:

2	5	3	0	4	1	2	4	3	5	2	3	2	2	4	1	3	2	3	4
5	3	2	2	3	3	2	3	4	5	1	2	2	4	2	2	3	2	1	2

- a) Group the data in a frequency table. Calculate the relative frequencies and the percentage frequency.
- b) Make a bar chart.
- c) Calculate the mean and the mode.

3. Calculate the mean mark for the marks grouped by intervals:

Mark	Range	Number of students
1	0-2	7
3	2-4	7
5	4-6	15
7	6-8	4
9	8-10	4

4. The Mathematics marks of a class in the first exam are as follows:

1'4	2'2	1'1	4'6	1'3	1'8	0'6
1'7	1'4	8'2	5'2	5'5	8'7	1
3'6	4	2'8	7'1	6	2'7	4'8
0'3	5'5	0'3	4'9	5'4	5'8	5'6
3'8	1	3	1'7	5'9	6'9	5'6

Calculate the mean mark for the marks grouped by intervals.

2. **CHARTS**

5. The marks obtained by a group of students in a test are:
15, 20, 15, 18, 22, 13, 13, 16, 15, 19, 18, 15, 16, 20, 16, 15, 18, 16, 14, 13.
a) Make a frequency distribution table for the data: absolute, relative and percentage.
b) Draw the corresponding frequency polygon.
6. The kind of films preferred by 80 young people are classified by the following table:

Genre	n_i	f_i	%	
Science			15%	
Comedy			25%	
Horror			10%	
Adventure			50%	

- a) Classify the studied character.
- b) Complete the gaps of the table.
- c) Work out the mode
- d) Draw a sector diagram.

3. MEASURES OF CENTRAL TENDENCY

7. The students of a school are consulted about the most popular sports. It got the following results:

<i>Sport</i>	<i>Frequency</i>
Athletics	20
Basketball	30
Football	40
Swimming	10

- a) Represent the data in a pie chart.
 - b) Complete the chart with the relative frequencies and percentage frequencies.
 - c) Calculate the central tendency measures.
8. The number of absent days to a training session for the members of a football team in a month were the following:

<i>Number of days</i>	0	1	2	3	4
<i>Frequency</i>	10	6	5	2	2

- a) Classify the studying character.
 - b) Represent the data in a bar chart.
 - c) Calculate all the central tendency parameters.
9. The chart collects the ages of the members of a Hiking Mountain Club
- a) Do a frequency table grouping data into 5 intervals of 5 years of amplitude starting at 10 and finishing at 35.
 - b) Work out the relative frequency and the percentage frequency for each.
 - c) Make the corresponding histogram.
 - d) Work out the central tendency measurements from the table on a)

10	21	31	11	24
16	26	16	26	30
24	22	32	19	20
18	34	12	21	27
25	23	18	19	23
27	12	28	20	15
16	24	14	27	14

10. The following table gathers the number of hours that a group of students spend weekly to do housework.

<i>Number of hours</i>	<i>Frequency</i>
0-2	2
2-4	5
4-6	6
6-8	10

- a. Classify the studying character.
- b. Calculate relative and percentage frequencies.
- c. Calculate the average.

- 11.** The weights of 50 people are represented by the following table:

Weight	x_i	n_i			
[60, 70]		6			
[70, 80]		20			
[80,90]		14			
[90, 100]		10			

- a) Complete the frequency table with relative and percentage frequency.
b) Work out the central tendency parameters.
c) Plot the histogram and frequency polygon.
- 12.** Find the mean, median and mode for the following set of numbers: 3, 5, 2, 6, 5, 9, 5, 2, 8, 6.
- 13.** There are the following values: 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4.
a) Make a frequency table: absolute, relative and percentage.
b) Calculate the mean, median and mode.
- 14.** 40 students in a class have obtained the following test scores out of 50.
3, 15, 24, 28, 33, 35, 38, 42, 23, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.
a) Classify the data by intervals from 0 to 50. Take as length 10 for each.
b) Make the absolute frequency table.
c) Work out the mode, mean and median.

- 15.** The math marks of a class of 25 students in the first exam are the following:

8	7	5	3	2	1	3	3	4	10	9	8	7	6	6	5	7	7	8	1	0	3	4	4	5
---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

1. Group the data in a frequency table, relatives and in percentage.
2. Calculate the mean.
3. Make a pie chart with two data: fail and pass.
- 16.** The mathematics marks in the first exam of three classes are as follows:

1º A:

3	0	0	6	4	0	2	4	7	3	4
4	8	5	6	2	1	2	7	0	5	9
0	5	2	1	5	2	7	8	2	3	0

1º B:

6	4	7	5	0	7	2
5	3	1	6	2	1	3
6	5	4	5	2	3	5
2	3	5	8	7	7	4
8	9	5	3	7	3	2

1º C:

5	6	5	2	5	2	5
7	7	2	6	5	4	5
4	1	6	0	2	3	5
7	4	7	3	0	0	5
6	5	2	0	3	7	3

1. Group the data in a frequency table, relatives and in percentages.
2. Make a bar chart per class.
3. Calculate the mean per class.
4. Order the groups according to their results.